Journal of Al-Qadisiyah for Computer Science and Mathematics Vol.17.(1) 2025,pp.Comp 103–121

Available online at www.qu.edu.iq/journalcm

JOURNAL OF AL-QADISIYAH FOR COMPUTER SCIENCE AND MATHEMATICS

ISSN:2521-3504(online) ISSN:2074-0204(print)

Analysis of AI- Empower Predictive Models for Predicting Student Performance in Higher Education

Husam Kadhim Gharkan [a,*b], *Mustafa Jawad Radif* [a], *Ali Hakem Alsaeedi* [a]

[a] *College of Computer Sciences and Information Technology, University of Al-Qadisiyah, Al-Qadisiyah, Iraq.*

[b] *Ministry of Education, Qadisiyah Education Directorate*

Email: cm.post23.4@qu.edu.iq, mustafa.radif@qu.edu.iq, ali.alsaeedi@qu.edu.iq

ARTICLE INFO

Article history:
Received: 25 /12/2024
Rrevised form: 12 /1/2025
Accepted : 9 /2/2025
Available online: 30 /3/2025

Keywords:

Predictive analytics,
Student performance,
Ensemble models,
Classification & Regression,
Deep learning,

ABSTRACT

This research presents a study and review of previous research. It demonstrates the use of the most important techniques in predictive analytics and machine learning algorithms to analyze historical data and accurately predict future outcomes of student performance. This research focuses on specific objectives, including techniques used to identify students at risk of poor academic performance or dropout and enable timely interventions to improve outcomes. Moreover, in this paper, the answers to questions, such as their benefits and limitations. By using data from sources such as academic records, attendance, and engagement metrics, educational institutions can uncover patterns in student behavior and performance. The research also presents the most important findings that were reached. The results show that predictive analytics not only improves individual student performance but also enhances the effectiveness of the institution by promoting a supportive and proactive learning environment. This approach provides educators and educational institutions with actionable insights to effectively enhance student retention and enhance academic achievement.

https://doi.org/10.29304/jqcsm.2025.17.11967

1. Introduction

In light of the scientific race that most educational institutions around the world [1, 2], predictive analytics is an important concept that will enhance the educational process [3-8]. Predictive analytics is a very powerful tool that uses data sets and artificial intelligence algorithms to identify patterns and behaviors to predict future outcomes for students in educational settings [9, 10]. Therefore, predictive analytics is crucial in helping educational institutions better understand student performance and behavior. This is done by analyzing various data, such as grades, attendance, and engagement levels. Predictive analytics can also provide future insights[11-13]. It also allows teachers to anticipate which students may be at risk of falling behind or dropping out.

This proactive approach enables schools and universities to intervene early and provide the necessary support for at-risk students' needs before encountering these problems [14-17]. For example, suppose there is a predictive model that tells us that a student is likely to struggle based on their current performance or engagement. In that

*Corresponding author: Husam Kadhim Gharkan

Email addresses: cm.post23.4@qu.edu.iq

Communicated by 'sub etitor'

case, teachers can provide additional tutoring, counseling, or other forms of assistance [18]. This predictive model helps proactively address the student's problems.

Predictive analytics not only helps improve individual student outcomes but also enhances the overall learning environment and how to make strategic decisions that will enhance educational processes [19-21]. Therefore, predictive analytics improves the approach of educational institutions, making it possible to design immediate and future interventions to meet the needs of students and the institution effectively, as well as provide counseling and advice by leveraging data-driven insights.

2. Mechanisms for Student Performance analysis

The literature review can be divided into several main sections based on the type of techniques used to predict students' future performance, which can be presented as follows:

2.1. Machine learning models (Classification & Regression)

This section provides a detailed analysis of various machine-learning models and techniques used to predict student performance. It will present the working and results of several algorithms, such as Random Forest (RF), XGBoost, Support Vector Machines (SVM), Neural Networks (e.g., MLP, RBFNN), Decision Tree (DT), K-Nearest Neighbors (KNN), and Fuzzy Logic techniques.

Alok, [22] proposes using MLP, AdaBoost, and XGBoost models as a solution to predicting student performance, The study utilizes a dataset containing student grades, demographic information, social aspects, and school-related features. The primary objective is to evaluate and compare the performance of these machine learning models in classifying student performance like "Good," "Fair," and "Poor.", The study's findings indicate that XGBoost outperforms both MLP and AdaBoost in predicting student performance, The findings are based on specific datasets related to student performance in mathematics and Portuguese language.

Singh, et al., [23] explore the application of machine learning (Logistic regression, random forest, gradient boosting regression, support vector machines) and fuzzy logic techniques to predict student performance. The study utilizes a specific dataset containing student scores in various subjects, class conduct ratings, and attendance records. The primary goal is to develop predictive models for student performance, including pass/fail outcomes, dropout identification, performance categorization, and Grade Point Average (GPA) prediction. Random Forest predicted the "Good" category well, and SVM had a high overall accuracy. However, the study lacks explicit handling of missing values, outliers, and feature selection.

Zhidkikh et al., [24] The research aims to address two main challenges. First, it seeks to validate the findings of a previous study by Van Petegem et al. (2022). Second, it investigates whether incorporating self-reported aptitude measures into the model can improve the accuracy of dropout prediction. The results show that the original approach was successfully reproduced. Adding self-report data improved accuracy in the first four weeks. The study uses many features without applying feature selection techniques.

Parkavi, et al.,[24] explore the use of Exploratory Data Analysis (EDA) and machine learning techniques, specifically K-Nearest Neighbors (KNN) and Multiple Linear Regression, to predict student performance. The study utilizes a specific dataset collected from 400 engineering students. The primary goal is to visualize student abilities and predict their performance, focusing on understanding the impact of the transition from traditional classrooms to online learning.

Sabbir, et al., [25] apply and evaluate the performance of five predictive models (SMOTE, XGBoost, KNN, Decision Tree, Random Forest), for enhancing the accuracy and effectiveness of student dropout prediction. A higher education student dataset from Kaggle containing student information, academic paths, and socio-economic factors was used. Random Forest and XGBoost emerged as the top-performing models. The authors suggest exploring inconsistent features and hybrid algorithms for further improvement.

Alamgir, et al., [26] employ machine learning models (Linear Regression, Random Forest, Neural Network (Multilayer Perceptron)). The study used a dataset that collected from year 2001 to 2015 from a local university, to predicting the future performance of undergraduate students and identifying at-risk students early on. The study explores that a Historical data and performance in core courses are significant predictors

Ramirez, [27] Develop and evaluate machine learning models (Logistic Regression, Random Forest, Gradient Boosting, Neural Networks, SVM) to predict student outcomes and identify key factors influencing success. The study includes collecting data from 15,000 undergraduate students from a large public university in the United States. Performance prediction accuracy up to 85% and Dropout prediction accuracy up to 85.6%.

Chukwuemeka, et al., [28] employ machine learning algorithms (linear regression, decision tree, Naive Bayes, KNN, K-means), to predict student academic performance and understand factors influencing it. The study utilizes a dataset collected from 1000 students from schools in Nigeria, the primary goal is to identify the factors that influence student evaluation and develop predictive models to classify students into different performance categories. The study discovered that the attributes that impact student evaluation are their ethnicity and parents' education level.

Tastanova, et al., [29] investigate the application of collaborative filtering algorithms in knowledge management to enhance predictive analytics in education. The study focuses on predicting student grades in specific courses based on their performance in previous courses. The primary goal is to evaluate the effectiveness of collaborative filtering techniques, specifically user-based collaborative filtering (Pearson correlation, Cosine similarity, Euclidean distance), in predicting student grades and compare its performance with other similarity measures. The dataset was collected from 250 students in Grades 10, 11, and 12 between 2021 and 2023 from Nazarbayev Intellectual school (NIS) in Shymkent.

Doctor, [30] proposes a solution that leverages data mining techniques and the Decision Tree algorithm to develop a predictive model to identifying students who may be at risk of failing their courses early in the semester. The model aims to analyze student performance data and identify patterns that can be used to predict the likelihood of a student passing or failing a course. By identifying at-risk students early on. The dataset comprises only 82 students. The predictive model achieved an accuracy of 0.7619, precision of 0.8333, recall of 0.8823, and F1 score of 0.8571.

Albahli, [31] proposes a method that combines the Synthetic Minority Oversampling Technique (SMOTE) to handle imbalanced data and Bayesian Optimization for hyperparameter tuning. The research uses a dataset of 5000 student records collected from five semesters at different universities in Saudi Arabia. The study uses Random Forest and Decision Tree as the machine learning algorithms for classification. Decision Tree with SMOTE and Bayesian Optimization outperformed other models.

Ni, et al., [32] present a novel approach combining Signed Graph Neural Networks (SGNNs) and Large Language Model (LLM) embeddings, which enhances the Predicting student performance on learner-sourced questions, particularly in cold start scenarios with limited data. The dataset includes Five real-world datasets from the PeerWise platform (biology, law, cardiff20102, sydney19351, sydney23146). The study shows the proposed LLM-SBCL model outperforms baseline methods.

Ramaswami, et al., [33] propose to use 'Anchors' for model explainability and 'what-if' counterfactuals for generating prescriptive analytics. the study aims to provide students with understandable reasons behind their predicted academic performance and offer actionable recommendations for improvement. The Dataset extracted from courses at an Australasian higher education institution. The study suggests that this approach can lead to more effective learning changes and improved student outcomes.

Gaftandzhieva, et al., [34] explores the use of statistical and machine learning techniques to predict students' final grades based on their online activities in a Moodle Learning Management System (LMS) and attendance in online lectures conducted via Zoom. The dataset included 105 students studying Object-Oriented Programming at the University of Plovdiv during the 2021-2022 academic year. The study used machine learning algorithm for prediction (RF, XGBoost, KNN, SVM). The Random Forest algorithm achieved the highest accuracy (78%).

Yagci, [35] investigates the use of machine learning algorithms to predict the final exam grades of undergraduate students based on their midterm exam grades, faculty, and department. The study utilizes a dataset of 1854 students who took the Turkish Language-I course at a state university in Turkey. The researcher employed six machine learning algorithms (RF, Neural Networks (NN), SVM, LR, NB, KNN). The performance of these algorithms was evaluated using some metrics such as classification accuracy (CA), precision, recall, F1-score, and area under the ROC curve (AUC). Achieved a classification accuracy of 70-75%.

The study by **Roslan, and Chen,** [36] investigates the application of data mining techniques to predict secondary school students' performance in English and Mathematics. The research aims to identify key predictors of student performance and classify students based on their performance levels, Primarily, the authors utilize Decision Trees (DT) and Naive Bayes (NB) for their analysis, noting that DT is particularly effective due to its interpretability and accuracy. However, the authors acknowledge limitations, including a small sample size and a lack of diversity, which may affect the generalizability of their results.

Khor, [37] explores data mining and machine learning techniques to predict academic performance in Massive Open Online Courses (MOOCs), focusing on the early identification of low-performing students. This proactive approach aims to facilitate timely interventions to improve learning outcomes and reduce dropout rates. Using a dataset that encompasses student demographics, academic history, and interactions within the Virtual Learning Environment (VLE), He applied three machine learning algorithms (decision trees, logistic regression, and neural networks) to develop predictive models. Decision trees algorithm outperformed others (70.27% accuracy).

Bressane, et al., [38] propose a fuzzy AI-based model using a fuzzy inference system (FIS) to forecast student performance and retention risk in engineering education. The methodology involves data collection through an online survey and historical records, statistical analysis to identify significant learning strategies using Wang & Mendel algorithm for fuzzy inference system modeling, along with other AI methods for comparison (CCN, MLP, GMDH, PNN, RBFN, GEP, Tree Boost, SVM), applied on dataset contain from 111 students. The fuzzy AI-based model outperformed other AI methods, achieving 94.0% accuracy during training and 91.9% generalization capacity over the testing dataset.

Adnan, et al., [39] employ various machine learning and deep learning algorithms (Random Forest, SVM, KNN, Extra Trees, AdaBoost, Gradient Boosting, DFFNN) to develop predictive models that identify at-risk students early in online courses for timely intervention. By using the Open University Learning Analytics Dataset (OULAD) The study highlights the significance of students' assessment scores, engagement intensity (clickstream data), and time-dependent variables in predicting their performance in online learning. The Random Forest algorithm showed progressive accuracy from 60%-91%.

Bujang, et al., [40] investigates the use of machine learning techniques for predicting student grades, with a focus on addressing the challenges posed by imbalanced multi-class datasets. The study employs various machine learning algorithms and techniques to develop a predictive model that can accurately classify students into different grade categories (Decision Tree J48, SVM, NB, KNN, LR, RF). The authors propose using the Synthetic Minority Oversampling Technique (SMOTE) and feature selection methods to address imbalanced dataset.

Kabathova and Drlik, [41] investigates the prediction of student dropout in university courses using machine learning techniques. The study emphasizes the importance of careful data preparation and feature selection when working with limited educational data. The authors compare the performance of several machine learning classifiers, including Logistic Regression, Decision Tree, Naive Bayes, Support Vector Machines, Random Forest, and Neural Network, on a dataset collected from an e-learning course over four academic years.

Yakubu, and Abubakar, [42] explores the use of machine learning, specifically logistic regression, to predict student academic success in a Nigerian university. The study utilizes enrollment data from the student information system, focusing on factors such as entry age, gender, state of origin, JAMB score (a standardized entrance exam), and level of study from a single institution in Nigeria. The primary goal is to identify the variables that significantly influence academic success. The study notes the exclusion of first-year students due to the absence of CGPA data, which might introduce bias.

Malini and Kalpana, [43] utilizes educational data mining techniques, employing classifiers like MLP, Bagging, and Boosting, to analyze student performance data and predict academic outcomes based on various attributes. The study aims to identify factors influencing student performance, particularly focusing on the impact of economic background. The study utilizes a dataset from the UCI repository, which contains information about secondary school students' performance in two subjects (Mathematics and Portuguese language). The results indicate that economic background significantly impacts student performance, with accuracy rates exceeding 70% for predicting performance based on economic attributes alone.

Bertolini, et al., [44] investigates the impact of incorporating feature selection techniques into data pipelines to forecast student success in an introductory biology course. The authors employ various feature selection techniques and machine learning models (Algorithms: LR, GLMNET, RF, XGBoost; feature selection techniques: CAE, FSA, IG, RAE) to assess their efficacy in predicting student outcomes. The study focuses on a specific context—an introductory biology course at a single institution.

El Guabassi, et al., [45] built a predictive model using different supervised machine learning algorithms to identify "weak" students who need extra help to improve their performance, identify the best machine learning algorithm to predict students' academic performance and understand the factors that affect students' academic success. Several supervised machine learning algorithms were applied to the Student Academic Performance Dataset (xAPI-Edu-Data), which includes 480 students with 19 features (ANOVA, Logistic Regression, Support Vector Regression, Logarithmic Linear Regression, Decision Tree Regression, Random Forest Regression, Partial Least Squares Regression). The algorithms were compared and evaluated using metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2).

Ashfaq et al., [34] used machine learning techniques to predict student performance. The study focused on the process of processing unbalanced data using (SMOTE and ADASYN), as well as feature selection using (FCBF and RFE). After balancing the data and selecting the features, the random forest achieved an accuracy of (86%).

Hasan, R., et al., [46] conducted a study to predict student performance using video learning analytics and data mining techniques. The study utilized data from the student information system, learning management system, and a mobile application to analyze student interactions with video lectures and other learning materials. The researchers employed eight different classification algorithms, including Random Forest, Logistic Regression, Decision Tree, k-nearest Neighbors, Naive Bayes, Neural Network, SVM, and CN2 Rule Induction. Data transformation and preprocessing techniques were applied to reduce the number of features and improve the models' performance. Random Forest achieved an accuracy (88.3%) in predicting successful students.

Bravo-Agapito et al. [35] focus on the early prediction of undergraduate student academic performance in a completely online learning environment. The study utilizes Moodle interaction data, student characteristics, and grades from 802 undergraduate students at a completely online university. The researchers employed exploratory factor analysis, multiple linear regressions, and cluster analysis to identify the variables that influence student performance and develop predictive models. The study identifies four key factors influencing academic performance: Access((variables related to students' interactions with the Moodle platform)), Questionnaire (number of questions answered and reviews of questionnaires), Task (number of assignments submitted, consulted, and completed), and Age(older students tend to perform slightly worse than younger ones).

Hashim et al. [36] focus on predicting student performance in their final examinations using supervised machine learning algorithms. The study utilizes a dataset from the College of Computer Science and Information Technology at the University of Basrah, containing student information such as study year, gender, birth year, registration status, employment, activity points, examination points, and final grade. The primary goal is to compare the performance of various supervised machine learning algorithms in predicting student outcomes, specifically their final grades and pass/fail status. The logistic Regression classifier achieved the highest accuracy in predicting the final grades of students (68.7% for passed and 88.8% for failed).

Rayasam, [47] explores the application of artificial intelligence to identify at-risk students in higher education, particularly those who may be suffering from mental illness but remain undetected. The study emphasizes the importance of proactive care and early intervention to prevent dropouts, leaves of, and even deaths related to mental health issues. The author proposes using non-invasive institutional data and machine learning to predict at-risk students before any symptoms manifest in 131 instances. The logistic regression model achieved an accuracy of 96.5% and an F1 score of 69.3% in predicting student dropout.

Yılmaz and Sekeroglu, [48] utilize the application of artificial intelligence (AI) techniques to classify student performance based on various indicators derived from a questionnaire. The study emphasizes the importance of predicting student outcomes to enhance educational quality. The research aims to classify students' final grades using AI, particularly focusing on personal information, educational preferences, and family background. Applying machine learning models (BPNN, RBFNN, Decision Tree, Logistic Regression). The dataset includes information from 101 students across three courses at Near East University. RBFNN outperforms other algorithms, achieving accuracy rates between 70% and 88%.

Table 1: Summarizes of Classification & Regression models research with the results.

Ref	Year	Problem of research	Solution (Method)	Accuracy	Advantage	Limitation
[9]	2024	Evaluating machine learning methods for predicting student outcomes.	XGBoost, MLP, ADA Boost.	97% (XGBoost), 95% (MLP), 81% (ADA Boost)	high performance for (XGBoost)	focusing on preprocessing only
[10]	2024	Analyzing student performance using machine learning and fuzzy logic.	Random forest, logistic regression, fuzzy logic, SVM.	LR:0.84 RF:0.94, SVM :0.98	Flexible prediction models, inclusion of fuzzy logic.	Fuzzy logic challenges in large datasets.
[11]	2024	Developing generalizable models for predicting student performance.	dropout prediction with combined self-reports	from 65% to 92%	generalizability in analytics.	Dependence on self-reported data quality.
[12]	2024	Predicting student performance using EDA and ML techniques.	KNN, multiple regression, EDA.	99% (multiple regression)	Sustainable education goals, effective EDA.	Limited to specific geographic data.
[13]	2024	Using predictive analytics to reduce dropout rates.	Random forest, XGBoost, KNN, Decision Tree.	0.99 AUC (Random Forest)	Addressed class imbalance, strong AUC values.	Class imbalance addressed, but dataset diversity limited.
[14]	2024	Identifying critical features for undergraduate computer science student success.	Random forest, Neural networks, Linear regression.	92%	Explored long-term data impact on predictions.	Focus on CS relative grading, less generalizability.
[15]	2023	Predicting student performance and dropout rates in higher education.	Logistic regression, random forests, neural networks.	0.85	Early dropout interventions,	Limited to one university dataset.
[16]	2023	Early prediction of low-performing students.	Linear regression, decision tree, Naïve Bayes, KNN, KMeans.	SVM : 95%	Improved understanding of academic behaviors.	Limited to specific student engagement factors.
[17]	2023	Enhancing the potential of artificial intelligence in education.	Collaborative filtering, AI-enhanced teaching frameworks.	-	Better cognitive and task organization skills in students.	Focuses only on AI-enhanced teaching effects.
[18]	2023	Predicting the probability of student success in courses.	Decision tree classification using CRISP-DM methodology.	0.7619	Early predictions help educators guide at-risk students.	Dependent on quality of integrated academic systems.
[19]	2023	Improving prediction accuracy using Bayesian optimization for imbalanced data.	Bayesian Optimization, SMOTE, Random Forest, Decision Tree.	Random Search: 95 % Improved with Bayesian Optimization	Enhanced handling of imbalanced data with SMOTE.	Time-consuming hyperparameter tuning process.

[20]	2024	Addressing noise issues in prediction of learner-derived data.	SGNNs, LLM embeddings, contrastive learning.	90.8%	Addresses noise and cold start problems effectively.	Relies on learnersourced data for cold start solution.

2.2. Ensemble models

Ensemble models are powerful for predicting student performance, combining techniques like Random Forest, Gradient Boosting, AdaBoost, and Stacking to enhance accuracy. Ensemble methods consistently outperform individual models, as these models use student engagement, demographic, and academic data to make robust predictions.

Alzahrani, [49] proposed the use of ensemble models (Random Forest, Gradient Boosting, Extra Trees, AdaBoost, Bagging, and a proposed Stacking model) and learning analytics techniques to predict student performance. The study uses the Open University Learning Analytics Dataset across different classification scenarios (4, 3, or 2 classes), and uses one dataset that may be which may limit the generalizability of the findings to other educational contexts., the study mentioned that the Stacking model emerged as the best performer, The models in this study were not compared with other advanced techniques like neural networks.

Al-Ameri, et al., [50] propose a novel approach that combines Convolutional Neural Networks (CNNs) for feature extraction with ensemble learning models, specifically an ensemble of Random Forest (RF) and Support Vector Machine (SVM). The CNN extracts intricate features from the multimedia data, capturing nuanced patterns and relationships. These extracted features are then used to train the ensemble model by using a dataset collected from the online learning platform MOODLE, containing student video interaction, academic information, and activity data. The Proposed ensemble model achieves 97.88% accuracy in predicting student academic success.

Nafea, et al., [51] proposed an ensemble model using a stacking classifier that combines the predictions of four base models, Random Forest, AdaBoost, Decision Tree, and Support Vector Machine. Logistic regression is used as the final estimator to produce the final prediction for student performance to enable early interventions and improve educational outcomes. The proposed ensemble model achieved an accuracy rate of 95%.that used the Open University Learning Analytics Dataset (OULAD) for predictive.

Balcioglu and Artar, [52] proposed applying machine learning and deep learning models (Decision Tree, Support Vector Machine, Neural Network, and Ensemble Model) for early prediction of student academic performance in higher education using the Open University Learning Analytics (OULA) dataset, which includes demographic information, prior academic history, and behavioral data of 173,913 students. The results showed that the ensemble Model outperforms other models in predicting student performance categories (Pass/Fail, Close to Fail, Close to Pass).

Alsulami, et al., [42] proposes a model that combines traditional data mining techniques (Decision Tree, Naive Bayes, and Random Forest) with ensemble methods (Bagging and Boosting) to enhance the prediction accuracy of student performance. The model also incorporates a voting process to refine predictions further. The dataset contains 480 records, and 17 attributes collected from the Kalboard 360 E-Learning system. Boosting with Decision Tree achieved the highest accuracy (77.9%).

Table 2: Summary of Ensemble models research with results

Ref	Year	Problem of research	Solution (Method)	Accuracy	Advantage	Limitation
[37]	2024	Limited accuracy of individual models in predicting student performance.	Ensemble models like stacking and bagging on OULAD dataset.	RF achieve 96% for 2 classes 86% for 3 classes 76% for 4 classes	Combines strengths of various models for better prediction.	Computational complexity increases with ensemble models.

[38]	2024	Underutilization of multimedia data from LMS for accurate predictions.	Convolutional Neural Networks with RF and SVM.	97.88%	Utilizes multimedia data for highly accurate predictions.	Reliance on LMS data for multimedia insights limits scope.
[39]	2023	Open issue of prediction accuracy in student performance classification.	Stacking ensemble using RF, DT, AdaBoost, SVM, Logistic Regression.	95%	Improves accuracy by reducing weaknesses of individual models.	Dependent on dataset quality and feature engineering.
[40]	2023	factors influencing academic performance.	Ensemble Model with SVM, NN, RF, and Decision Tree	87%	Actionable models and techniques to solve targeted problems.	Limited to OULAD dataset;
[42]	2023	Challenges in analyzing e-learning data and improving prediction accuracy.	Bagging, boosting, and voting on DT, Naive Bayes, RF.	77% (Boosting)	Enhances interpretability of educational data mining.	Specific to e-learning data; generalization not tested.

2.3. Deep learning

Deep learning techniques like LSTM, CNN, and DNN are effective for predicting student performance and leveraging engagement data, demographics, and online activity. Models integrating methods such as SPTLO+DLSTM, ANN-LSTM, and Sequential Engagement Prediction Networks (SEPN) achieved high accuracies by analyzing online patterns and optimizing feature selection.

Hussain, et al., [53] proposed a deep learning framework using Levenberg Marquardt Algorithm (MLA) to predict academic performance and aid decision-making. The study includes data preprocessing and cleaning for datasets collected from Govt. Post Graduate College Safdarabad. The predictive achieved a higher accuracy of 88.6%

Sharada, [54] proposes the SPTLO-based DLSTM model to predict student performance by leveraging knowledge tracing and student psychology. The authors highlight the importance of accurate student performance prediction in online learning environments to provide personalized and effective learning experiences. The methodology contains: (Data collection, Data normalization, Data augmentation (SMOTE), Feature fusion DMN with Ruzicka similarity, Deep Knowledge Tracing (DLSTM), and Optimization (SPTLO, combining SPBO and TLBO). SPTLO+DLSTM achieved accuracy (92.5%), MAE (0.064), and RMSE (0.312) compared to other models.

Sharma and Bhardwaj [55] utilize an LSTM-based model using clickstream data and demographic information to analyze students' online activities and engagement patterns. The study used the OULAD dataset with 170 features extracted from VLE log data and additional demographic information to Predict student learning abilities. The primary goal is to develop a predictive model that can identify at-risk students and forecast their academic outcomes based on their online behavior and engagement patterns. The model achieved 75% accuracy.

Al-Azazi and Ghurab, [55] proposed a novel deep learning model, ANN-LSTM, for the early prediction of student performance in Massive Open Online Courses (MOOCs). The model aims to classify student performance into multiple categories (distinction, pass, fail, and withdrawn) using demographic and clickstream data. The authors compare the performance of ANN-LSTM with two baseline models, Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU). By using dataset of Open University Learning Analytics Dataset (OULAD). ANN-LSTM achieved around 70% accuracy at the end of the third month, outperforming RNN (53%) and GRU (57%).

Chan, [56] propose ed to apply multimodal learning analytics (MMLA) using machine learning and DNN for Predicting behavior change in students with special education needs (SEN) during applied behavior analysis (ABA) therapies. Data was collected from 1,130 ABA therapy sessions with 32 SEN students by using an IoT-based system. RF classifier achieves an accuracy of 97.79%.

Shafiq, et al., [57] proposes a conceptual predictive analytics model to identify at-risk students in Virtual Learning Environments (VLEs) using machine learning techniques. The study focuses on the Open University (OU) dataset. It aims to evaluate the effectiveness of unsupervised machine learning approaches, specifically K-means clustering, in predicting student performance compared to supervised methods. The focus on K-Means clustering may overlook the potential benefits of hybrid models or other unsupervised methods that could provide richer insights into student behavior.

Hooda et al. [58] proposed applying a Fully Convolutional Network (FCN) to predict student academic performance, particularly in the context of online learning due to the COVID-19 pandemic. The study used an Open University Learning Analytics Dataset (OULAD). The proposed improved FCN model achieved an accuracy of 84%, outperforming the conventional ANN model. It also attained a Recall of 0.88, an F1-score of 0.91, and a precision of 0.93.

Tao, et al., [58] Propose a DNN-based prediction and early warning system with improved K-nearest neighbor clustering, to Predict unknown course grades and recommend optimal learning methods. The study used Students' performance data from the College of Metallurgical Engineering in Anhui Province, China University. the methodology of the study is to Compare linear regression, random forest, BPNN, and DNN, and propose DNN-based prediction and improved KNN clustering.

Liu, et al., [59] Proposed a student achievement prediction model based on evolutionary spiking neural networks, to improve the accuracy of student achievement prediction models. The study used the xAPI-Edu-Data dataset, 480 student records with 17 variables, and the student performance dataset from the UCI Machine Learning Repository, 395 student records with 33 variables. The proposed model outperforms other experimental models in classification accuracy on both datasets

Hernandez, et al., [60] explore the application of Artificial Neural Networks (ANNs) in predicting academic performance in higher education. The study employs a Multilayer Perceptron (MLP), a common type of ANN known for its predictive capabilities. The MLP consists of an input layer, a hidden layer, and an output layer. The study uses the online learning method with the backpropagation algorithm to train the MLP. The study used a dataset of 162,030 students from private and public universities in Colombia. ANNs achieved an accuracy of 82% for high performance and 71% for low performance. Outperformed other algorithms in recall and F1 score.

Neha et al. [54] used the Deep Neural Network (DNN) model to predict students' academic performance **accurately** and to enhance the prediction of students' academic performance through multiple predictive variables and internal and external factors that affect performance.

Nabil, et al., [61] utilize the application of deep learning, specifically Deep Neural Networks (DNNs), in predicting students' academic performance. The primary goal is to identify students at risk of failing, particularly in challenging courses like Programming and Data Structures. The study utilizes a commonly used educational dataset (xAPI-Edu-Data from Kalboard 360)., employing various machine learning models like DNN, decision tree, random forest, gradient boosting, logistic regression, support vector classifier, and K-nearest neighbor to predict student performance in upcoming courses based on their grades in previous courses. The research also addresses the issue of imbalanced datasets by employing resampling techniques like SMOTE, ADASYN, ROS, and SMOTE-ENN. The DNN achieved an accuracy of 89%, outperforming traditional models in terms of F1-score. Other models like SVC, KNN, and RF reached 91% accuracy but struggled with the minority class prediction.

Turabieh et al. [56] proposed an enhanced HHO algorithm for feature selection and KNN to evaluate the quality of selected features for Predicting student performance to improve educational outcomes and reduce failure rates. Several machine learning classifiers (kNN, LRNN, Naive Bayes, ANN) are used to assess the prediction system's performance. the study used A dataset from the UCI Machine Learning repository, which contains data about secondary school students in Portugal, including 32 features (demographic, social/emotional, school-related) and the final grade. The proposed method achieved a prediction accuracy of 90.16%.

Song et al. [57] proposed a Sequential Engagement Based Academic Performance Prediction Network (SEPN) model for predicting students' academic performance in online education based on their engagement patterns and other factors. The study methodology included transforming students' daily online activities into sequential engagement matrices, using CNN to detect engagement patterns, using LSTM to learn sequential interactions between engagement features and other factors, and combining classification and regression losses. The results show that the SEPN model outperformed traditional models in terms of prediction accuracy, recall, F1 score, and Mean Square Error (MSE).

Abubakari and Suprapto [58] explore the application of educational data mining to predict student academic performance. The authors utilize a neural network model to analyze student data and classify students into performance categories (Good, Average, Poor). The primary goal is to create a reliable predictive model that can assist in making informed decisions regarding student academic outcomes. The study uses a dataset containing demographic, socioeconomic, and academic information of 131 students and a neural network algorithm for classification. Adam's optimizer significantly improved accuracy to over 96%. SGD optimizer resulted in accuracy below 80%.

Waheed, et al., [62] investigate the performance of Long Short-term Memory (LSTM) in predicting students at risk of failing a course in self-paced online education. The study utilizes the Open University Learning Analytics Dataset (OULAD), which comprises data from 32,593 students, with 69% pass and 31% fail instances. The dataset includes student demographic data, clickstream data representing student engagement with the Moodle platform, and assessment-related information. The deep LSTM model outperforms all other algorithms in predicting student performance, achieving an accuracy of 84.57%.

Waheedet et al. [63] propose using deep learning models to predict student academic performance using data from Virtual Learning Environments (VLEs). The study focuses on predicting four categories of student performance: 'withdrawn-pass,' 'pass-fail,' 'distinction-pass,' and 'distinction-fail.' The authors utilize a deep artificial neural network (ANN) model and compare its performance with baseline models like logistic regression and support vector machine. The deep ANN model achieved a classification accuracy of 84-93%.

Raga, and Raga, [63] Develop a DNN model that utilizes online activity attributes (extracted from Moodle logs) to predict student outcomes at midterm and final periods, using activity data generated before the midterm. the study shows the Highest accuracy for predicting outcomes, 91.07% (COM course), and the highest accuracy for predicting midterm outcomes, 80.36% (COM course).

Tsiakmaki, et al., [64] propose a transfer learning methodology using deep neural networks to predict student performance in new courses by utilizing pre-trained models from related courses. The dataset included five datasets from compulsory undergraduate courses (three from Chemical Engineering, two from Physics) at Aristotle University of Thessaloniki. The results show that the transfer learning models DNN generally outperformed the baseline models.

Karlık, and Karlık, [65] proposed to compare five different neural network models to find the best one for predicting high school student performance. Survey data was collected from 11th-grade students at Taldykorgan Kazakh Turkish High School (Kazakhstan). The algorithms are Multi-Layered Perceptron (MLP), Fuzzy Clustering Neural Networks (FCNN), and Deep Neural Networks (DNN), and the results show that FCNN and DNN2 models provided the best accuracy at 91.67% and 91.877% respectively.

Hussain, et al., [66] employed deep learning and linear regression models to predict student performance based on their scores in major subjects. The dataset was collected from three colleges in Assam, India, and included 10,140 records of Bachelor of Arts students. The results show the Deep learning model outperforms linear regression, with Mean absolute error (MAE) for deep learning at 1.61 and MAE for linear regression at 1.97.

Table 3: Summary of Deep learning models research with results

Ref	Year	Problem of research	Solution (Method)	Accuracy	Advantage	Limitation
[41]	2024	Current methods do not have sufficient features	Levenberg Marquardt Algorithm (MLA), a neural network-based approach	88.60%	Improves early warning systems for at-risk students.	Small dataset limits generalizability.
[42]	2024	Low accuracy in predicting student performance.	Student Psychology-Based Optimization with Deep Long Short-Term Memory (DLSTM).	92.50%	Enhances predictive power of knowledge tracing models.	Complex model training

[43]	2023	Improving prediction of at-risk students using clickstream data from virtual learning environments.	Neural Network (FCNN) and Long Short-Term Memory (LSTM).	75%, Improved with processed clickstream	Analyzes student interaction data for actionable insights.	Dependent on quality and consistency of VLE logs.
[44]	2023	Lack of multi-class early prediction models for student performance	ANN-LSTM combining Artificial Neural Networks with Long Short-Term Memory.	70%	Early predictions	Relatively low accuracy compared to binary classification models.
[45]	2023	Enhancing behavioral prediction for students using multimodal learning analytics.	Multimodal data analytics with IoT sensors and wearable technology.	98%	Combines diverse data for highly accurate predictions.	Requires integration of multiple data streams.
[46]	2022	High dropout rates in VLE systems	K-Means clustering for unsupervised prediction of at-risk students.	Improved with K-Means clustering	Supports unsupervised analysis for early interventions.	Limited generalizability to other datasets.
[47]	2022	Low prediction accuracy in assessing student quality in higher education.	Fully Connected Network (FCN) algorithm enhanced with Learning Analytics.	84%	Combines EDM and LA	Dependent on OULAD dataset.
[48]	2022	Difficulty in early warning of grades	Deep Neural Networks combined with improved KNN clustering for early warning systems.	20% improvement	Early warning with improved clustering accuracy.	Computationally intensive for clustering large datasets.
[49]	2022	Challenges in predicting student achievement using conventional models.	Evolutionary spiking neural networks with membrane algorithms for hyperparameter optimization.	0.84375	Optimizes prediction using evolutionary learning.	Requires complex parameter tuning.
[50]	2021	Complexity in systematically implementing ANNs to predict academic performance.	Systematic ANN implementation for classifying academic performance.	82%	Systematic implementation for wider ANN adoption.	Limited to specific datasets.

2.4. Hybrid model

Hybrid models effectively predict student performance by combining clustering (e.g., K-means) and classification (e.g., Random Forest, SVM) techniques and achieve high accuracies by integrating feature selection and advanced algorithms.

Sultan Alalawi [65] proposes using the k-means clustering algorithm to analyze student performance data and improve educational strategies. Apply that to 49,588 student records in the Oman Education Portal. The methodology followed was data selection, data pre-processing, k-means cluster analysis, and data extraction. The results were Cluster 0: 4,621 students (9% - Failure), Cluster 1: 27,509 students (55% - Excellent), Cluster 2: 16,527 students (33% - Good) and Cluster 3: 931 students (2% - Average).

Nafuri et al. [67] propose a clustering-based approach to classify B40 students based on their academic performance and behavior in HEIs.to discover the reason for High dropout rates among B40 (40% of households with the lowest income) students in Malaysian higher education institutions (HEIs). The study employs three unsupervised models, K-means-BIRCH-DBSCAN, which are applied to records of 248,568 students with 53 attributes obtained from Malaysian higher education institutions (HEIs). KMoB demonstrated the highest performance among the models assessed, assisting in identifying factors influencing dropout rates.

Almasri, et al., [67] propose A unified framework for a novel supervised cluster-based (CB) classifier model that combines clustering and classification techniques to improve accuracy in predicting student performance. The study methodology included the clustering Phase (Historical student records are grouped into homogeneous clusters using clustering algorithms, classifier Model (For each cluster), and feature Selection (Relevant features are selected using wrapper-based techniques to enhance model accuracy. The dataset was collected from the Management Information Systems department of Balqa Applied University (BAU), Jordan. The CB classifier model demonstrated an accuracy (up to 96.96% with relevant features).

Omar, et al., [68] propose a combined methodology using the k-means clustering algorithm and the Elbow method to analyze and evaluate student performance accurately. the study Methodology was Data Preprocessing, Elbow Method (Determining the optimal number of clusters(k)), and K-means Clustering. The optimal number of clusters identified was k=3 after applying the Elbow method to the dataset from Oakland University, specifically focusing on computer science students.

Goh, et al., [69] proposed applying K-means clustering to analyze assessment scores and group students to identify potential health-related issues affecting their academic achievements. The dataset Comprises student assessment scores from 106 students across 8 subjects, including Mathematics, Biology, Malay Language, Additional Mathematics, Chemistry, English, History, and Physics. The results show that cluster performance (4, 5, and 6 clusters) varied, with students achieving overall performance above 70% (A category) and below 50% (F category).

Ahmed, [70] Proposed a model using K-means clustering and various classification algorithms (SVM, Decision Tree, Naive Bayes, KNN) to predict student performance and identify factors affecting it in e-learning environments. The methodology included data preprocessing, feature extraction using random forest, K-means clustering with Davies-Bouldin index, and hyperparameter tuning using grid search. The dataset was collected from students of Wollo University and Kombolcha Institute of Technology. SVM achieved accuracy (96%), Decision Tree (93.4%), and Naive Bayes y (83.3%).

Vimarsha, et al., [71] proposed a co-evolutionary hybrid intelligence (CHI) approach, which integrates human expertise with artificial intelligence to student performance prediction. The study methodology contains from two models. Module 1 for Classifies teachers into classes of expertise based on the accuracy of several machine learning algorithms. Module 2, proposes the RLCHI model, which combines reinforcement learning algorithms with insights from expert teachers to predict student performance. The RLCHI model achieved Accuracy 79.5%, Precision 49.9%, Recall 52.3%, and F-Score 53.2%.

Al-Tameemi, et al., [72] proposed a hybrid machine learning approach that integrates clustering and classification techniques to predict better student performance based on their learning behaviors. The study used the North American University (NAU) dataset and the Open University Learning Analytics Dataset. The methodology includes Data Collection, Data Pre-processing, Unsupervised Learning (K-Means Clustering, Principal Component Analysis (PCA)), and Supervised Learning (Random Forest (RF), Decision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Naive Bayes (NB) and Employed Feedforward Dense Networks for final predictions). The study found that the hybrid model combining FDN, RF, and DT outperformed other classifiers in predicting student academically.

Alshanqiti, and Namoun, [73] proposed a hybrid regression model combined with a multi-label classifier that predicts student performance and identifies influential factors contributing to that performance. Hybrid Regression Model (HRM) Combines Collaborative Filtering (Matrix Factorization), Fuzzy Set Rules, and Lasso Linear Regression to provide robust predictions of future student performance. Multi-Label Self-Organizing Map (MLSOM) is used to identify multiple factors influencing student performance and dataset preparation. The proposed hybrid model showed significant improvements in prediction accuracy compared to traditional single-baseline models, such as linear regression and matrix factorization.

Evangelista, [74] proposed a hybrid machine learning framework combining classification algorithms and ensemble methods to predict students' performance in VLEs. The research follows the Cross Industry Standard Process for Data Mining (CRISP-DM) model, incorporating Data Harvesting (Collecting data from VLE logs), Data Preprocessing, Feature Selection Using filter-based (CFS Subset Eval) and wrapper-based (Classifier Subset Eval) methods to select the most relevant features, and Modeling and Validation for training various classifiers and ensemble methods, employing 10-fold cross-validation for evaluation. The Boosting algorithm achieved an accuracy of 98.56%; the Bagging algorithm achieved an accuracy of 98.35%.

Allohibi, [75] Proposed a hybrid classifier (PHC) that integrates multiple classification algorithms to improve prediction accuracy for student performance. Classification Algorithms Employed Random Forest (RF), C4.5, CART, Support Vector Machines (SVM), Naive Bayes (NB), K-Nearest Neighbors (KNN), Proposed Hybrid Classifier (PHC) achieved an accuracy of 92.40.

Zhuang et al. [76] Proposed a Prediction and Alert model that integrates data from both online and offline learning to improve the accuracy of learning performance predictions for students. The methodology includes Data Collection (Collected from 50 eighth-grade students over a 16-week period), Data Analysis (Pearson correlation coefficients, ridge regression), and algorithm Development (Develops a predictive model to alert teachers about students' performance based on integrated data analysis). The study shows that a model demonstrated feasibility and effectiveness in predicting student learning performance.

Table 4: Summary of Hybrid model research with results

Ref	Year	Problem	Solution (Method)	Accuracy	Advantages	Limitations
[65]	2023	Lack of efficient data mining for student performance analysis in Oman.	K-Means clustering optimized with Elbow method for OEP data.	High with optimized K-Means	Effective clustering on large education datasets.	Focus on specific regional data.
[58]	2022	High dropout rates among B40 students in Malaysian universities.	K-Means, BIRCH, and DBSCAN clustering models optimized for B40 students.	99.81	Supports targeted interventions for B40 students.	Limited to B40 demographics.
[59]	2020	Difficulty in selecting optimal predictors for student performance.	Clustering-based EMT unified, clustering and classifiers.	96.25%	Combines clustering and classifiers for improved prediction.	Dependent on feature selection accuracy.
[60]	2020	Challenges in combining clustering techniques for better analysis.	K-Means and Elbow clustering techniques	Enhanced with Elbow method	Accurate analysis with multiple clustering techniques.	High computational cost.
[61]	2020	Health conditions linked to poor student performance within clusters.	K-Means clustering to academic performance.	87%	Health insights linked to academic performance.	Limited to health-related insights.
[62]	2022	High dropout rates among B40 students in Malaysian universities.	K-Means, BIRCH, and DBSCAN clustering models optimized for B40 students.	Improved with K-Means	Supports targeted interventions for B40 students.	Limited to B40 demographics.
[63]	2024	Lack of transparency in performance prediction models.	combined with teacher expertise and hybrid intelligence.	79.50%	Improved decision-making with hybrid intelligence.	Heterogeneity in datasets can cause bias.
[64]	2024	multi-class prediction using hybrid ML methods.	K-Means clustering, and supervised classifiers.	RF, DT, and FDN 93.8%, 90.7%, and 90.9%	Combines unsupervised and supervised methods.	Complexity in hybrid approach implementation.
[65]	2020	Difficulty in identifying the most influential factors in student performance.	Hybrid regression with Lasso, fuzzy collaborative filtering.	Improved with hybrid regression	Identifies factors impacting performance.	Requires extensive tuning for models.

[66]	2021	Lack of robust prediction models for virtual learning environments.	Hybrid ensemble methods (Bagging, Boosting, Voting) for VLE data.	98.15	Improved VLE predictions with ensembles.	Heavy reliance on ensemble techniques.
[67]	2024	Improving predictive accuracy by combining classifiers.	PHC combining Random Forest, CART, and C4.5 classifiers.	92.40%	Robust predictions with hybrid classifiers.	increases computational overhead when Combining classifiers
[76]	2020	Challenges in combining clustering techniques for better analysis.	K-Means and Elbow techniques applied to GPA and test scores.	From September to December 85% to 100 %	Accurate analysis with multiple clustering techniques.	High computational cost.

It becomes clear from reviewing the previous literature that the approach followed by researchers to reach the prediction processes of students' academic performance consisted of four basic branches, which are classification using different machine learning algorithms, using ensemble models, using deep learning, and using the hybrid model, as shown in Figure 1.

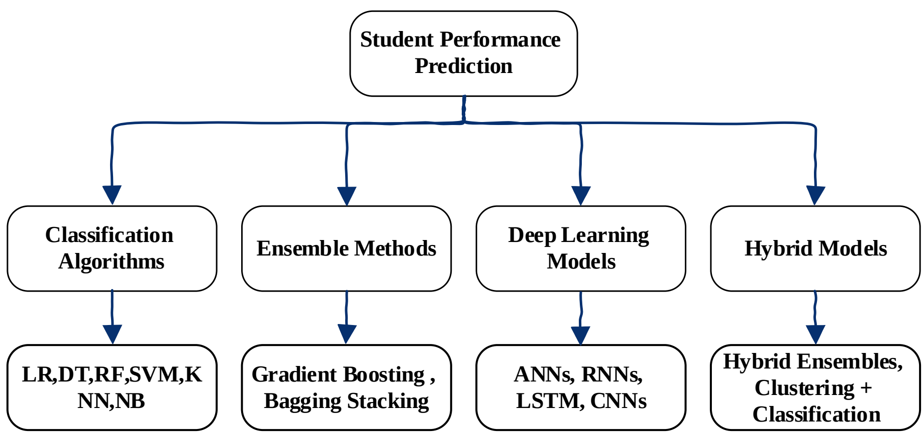

Figure (1) : General structure for using student performance prediction techniques

By reviewing the literature review mentioned above, it becomes clear to us that there are many researchers from different institutions who have conducted many studies on predicting student performance, including predicting future student performance, as well as predicting access and enrollment rates in education, etc. During the last five years alone, we see that approximately more than 70 important studies have been conducted using various techniques in the field of artificial intelligence. As shown in figure 2 and 3

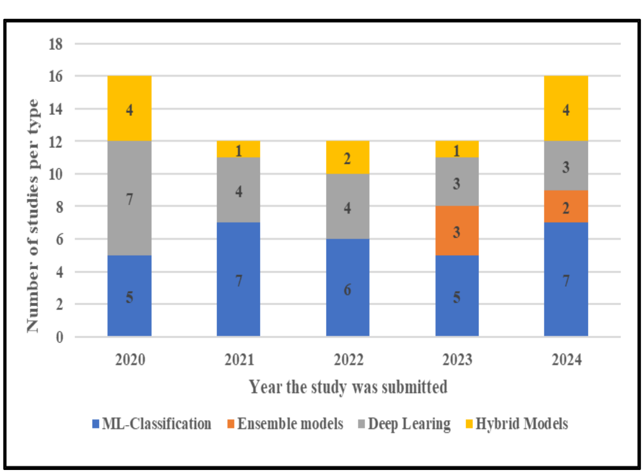

Figure (2) : Number of studies per type during the year.

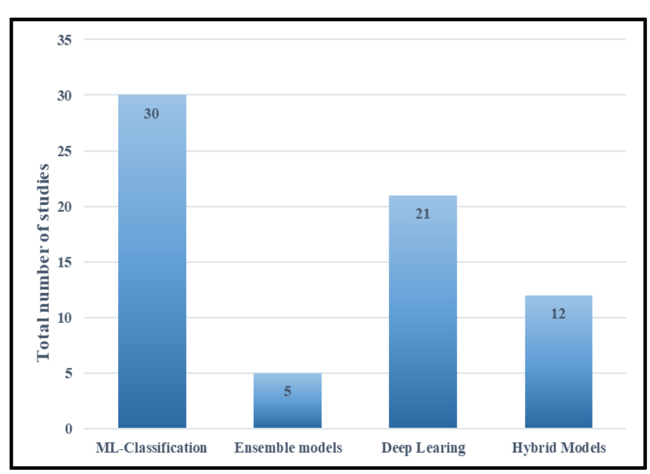

Figure (3) : Total number of studies during (2020-2024)

3. Datasets Description

In this section, we will discuss three important and most widely used databases in student performance that are available for scientific research. We will explore them and show their most important components.

3.1. Open University Learning Analytics dataset

The Open University Learning Analytics dataset (OULAD) contains data from courses at the Open University (OU) [76], a large distance-learning university. The dataset includes demographic data about students and clickstream data logging their interactions with the university's virtual learning environment (VLE). This allows analysis of student behavior and learning. The dataset covers 22 courses taken by 32,593 students, including their assessment results and 10,655,280 daily activity logs in the VLE. The dataset was anonymized to protect student privacy, with quasi-identifying attributes like gender, age, and location generalized using the ARX anonymization tool. The OULAD dataset is freely available for research purposes under a CC-BY 4.0 license and has been certified by the Open Data Institute. The dataset enables research on predictive models for student assessment and learning outcomes, as well as analysis of how the VLE structure impacts learning.

Screenshot (1): Sample of OULD dataset ,(studentInfo.csv)

code_module	code_presentation	id_student	gender	region	highest_education	imd_band	age_band	num_of_prev_attempts	studied_credits	disability	final_result
AAA	2013J	11391	M	East Anglian Region	HE Qualification	90-100%	55<=	0	240	N	Pass
AAA	2013J	28400	F	Scotland	HE Qualification	20-30%	35-55	0	60	N	Pass
AAA	2013J	30268	F	North Western Region	A Level or Equivalent	30-40%	35-55	0	60	Y	Withdrawn
AAA	2013J	31604	F	South East Region	A Level or Equivalent	50-60%	35-55	0	60	N	Pass
AAA	2013J	32885	F	West Midlands Region	Lower Than A Level	50-60%	0-35	0	60	N	Pass
AAA	2013J	38053	M	Wales	A Level or Equivalent	80-90%	35-55	0	60	N	Pass
AAA	2013J	45462	M	Scotland	HE Qualification	30-40%	0-35	0	60	N	Pass
AAA	2013J	45642	F	North Western Region	A Level or Equivalent	90-100%	0-35	0	120	N	Pass
AAA	2013J	52130	F	East Anglian Region	A Level or Equivalent	70-80%	0-35	0	90	N	Pass
AAA	2013J	53025	M	North Region	Post Graduate Qualification		55<=	0	60	N	Pass

3.2. Students' Academic Performance Dataset

The dataset, called xAPI-Edu-Data, was collected from a learning management system (LMS) called Kalboard 360[77]. It includes 480 student records with 16 features. Features fall into three categories: demographic, academic background, and behavioral. Demographic features Include gender and nationality. Students come from various countries, such as Kuwait, Jordan, Palestine, Iraq, and Lebanon. Academic Background Features Include educational stage (e.g., lower level, middle school, high school), Grade level (e.g., G-01, G-02), Section ID (classroom), Behavioral Features Includes raised hand in class, visited resources (course content), Viewing announcements. Participation in discussion groups. Parent participation (answering surveys, school satisfaction); students are classified based on absence days (above 7 or under 7)

Screenshot (2): Sample of Students' Academic Performance dataset (student-Math)

gender	NationalITy	PlaceofBirth	StageID	GradeID	SectionID	Topic	Semester	Relation	raisedhands	VisITedResources	AnnouncementsView	Discussion	ParentAnsweringSurvey
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	15	16	2	20	Yes
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	20	20	3	25	Yes
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	10	7	0	30	No
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	30	25	5	35	No
M	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	40	50	12	50	No
F	KW	KuwaIT	lowerlevel	G-04	A	IT	F	Father	42	30	13	70	Yes
M	KW	KuwaIT	MiddleSchoo	G-07	A	Math	F	Father	35	12	0	17	No

3.3. Student Performance Dataset (Student -Math and Student -Por Dataset)

The dataset focuses on student achievement in secondary education in two Portuguese schools. It includes student grades[78], and demographic, social, and school-related features; two datasets are provided for distinct subjects: Mathematics (mat) and Portuguese language (por). The target attribute is G3, representing the final year grade (issued in the 3rd period), G3 has a strong correlation with attributes G2 (second-period grade) and G1 (first-period grade). Features include information such as school type, student sex, age, family size, parental education, mother's and father's jobs, travel time, study time, and more, Quality of family relationships, free time, going out with friends, alcohol consumption, health status, and school absences are also part of the dataset.

Screenshot (3): Sample of Students Performance dataset (student-math) table)

school	sex	age	address	famsize	Pstatus	Medu	Fedu	Mjob	Fjob	reason	guardian	traveltime	studytime	failures	schoolsup	famsup	paid	activities	nursery	higher	internet	romantic
GP	F	18	U	GT3	A	4	4	at_home	teacher	course	mother	2	2	0	yes	no	no	no	yes	yes	no	no
GP	F	17	U	GT3	T	1	1	at_home	other	course	father	1	2	0	no	yes	no	no	no	yes	yes	no
GP	F	15	U	LE3	T	1	1	at_home	other	other	mother	1	2	3	yes	no	yes	no	yes	yes	yes	no
GP	F	15	U	GT3	T	4	2	health	services	home	mother	1	3	0	no	yes	yes	yes	yes	yes	yes	yes
GP	F	16	U	GT3	T	3	3	other	other	home	father	1	2	0	no	yes	yes	no	yes	yes	no	no
GP	M	16	U	LE3	T	4	3	services	other	reputation	mother	1	2	0	no	yes	yes	yes	yes	yes	yes	no
GP	M	16	U	LE3	T	2	2	other	other	home	mother	1	2	0	no	no	no	no	yes	yes	yes	no

Data Anonymization and Bias Audits

Data Anonymization : One of the important things that must be followed in the use of data is anonymization [79], the purpose of which is to protect the privacy of students and prevent unauthorized identification. The most important methods used for De-identification. This is done by removing personally identifiable information (PII), such as names or student IDs. Anonymization technology ensures that each record in a data set cannot be uniquely identified by making it indistinguishable from other records based on at least a set of quasi-identifiable identifiers (QIs). Quasi-identifiable identifiers are attributes that, when combined, can identify an individual. Examples include age, zip code, or gender.

Bias Audits: The purpose of these checks is to identify and address potential biases in the data [80], such as students who are overrepresented in the data with certain characteristics, or in predictive models that may be disadvantageous to certain student groups. This is done by analyzing the data set and reviewing the representativeness of the data to ensure that it does not exclude or disproportionately represent certain demographics (e.g., gender, race, socioeconomic status). Measures of model fairness can also be used by evaluating models using fairness criteria such as demographic equivalence, equal opportunity, or unequal impact. Testing predictions to identify patterns of bias or systematic discrimination.

Discussion

This review shows that AI-powered predictive analytics can significantly improve student performance in higher education. By using data and advanced machine learning models, schools and universities can predict which students might struggle and take steps to support them early. Tools like Random Forest, neural networks, and hybrid models are some of the most accurate and useful approaches. They help educators understand patterns in student performance, such as how attendance, grades, and engagement affect outcomes.

One of the most important benefits of predictive analytics is its ability to identify students at risk of dropping out or failing. Early warning systems can alert teachers, allowing them to offer extra help like tutoring or counseling before problems become too big. This personalized approach helps students stay on track and improves their chances of success. At the same time, these tools also help institutions manage resources better and improve overall results.

However, challenges remain. For example, the data used in predictive models is not always perfect. Missing or inaccurate information can reduce the reliability of predictions. Additionally, there are concerns about fairness and ethics. If the data used to train the models is biased, the predictions might unfairly disadvantage some students. Privacy is another issue, as schools must ensure student information is handled responsibly.

Another challenge is that many predictive models work well for specific groups of students or certain types of schools but may not apply to all. This means researchers need to develop tools that work for various students and educational settings. For non-experts, many models are also difficult to understand, making it hard for teachers to trust the results or use them effectively. Clearer and simpler tools are needed to address this issue.

Finally, combining predictive analytics with other educational tools, like learning management systems, could make them even more powerful. This would allow schools not only to predict student performance but also adjust teaching strategies in real time to meet individual needs.

Conclusion

AI-powered predictive analytics transforms higher education by enabling institutions to make data-driven decisions that enhance student performance. This systematic literature review has demonstrated the efficacy of various machine learning models, such as Random Forest, neural networks, and hybrid approaches, in predicting academic success and identifying at-risk students. These advancements pave the way for more personalized, inclusive, and effective educational practices. Even though these tools are promising, some areas still need improvement. Better data, fairer algorithms, and more user-friendly systems are essential for making predictive analytics useful for everyone. Future research should focus on creating tools in different educational settings and exploring how these technologies can work together with other tools like adaptive learning systems. By addressing these challenges, AI-powered analytics can help educators provide the support students need, reduce dropouts, and improve learning outcomes. These advancements not only help individuals but also contribute to creating more inclusive and successful educational systems. Even though these tools are promising, some areas still need improvement. Better data, fairer algorithms, and more user-friendly systems are essential for making predictive analytics useful for everyone. Future research should focus on creating tools in different educational settings and exploring how these technologies can work together with other tools like adaptive learning systems.

References

[1] J. Douglass, "The global higher education race," *International Higher Education*, no. 49, 2007.

[2] A. H. Alsaeedi, D. Al-Shammary, S. M. Hadi, K. Ahmed, A. Ibaida, and N. AlKhazraji, "A proactive grey wolf optimization for improving bioinformatic systems with high dimensional data," *International Journal of Information Technology*, vol. 16, no. 8, pp. 4797-4814, 2024.

[3] W. W. Eckerson, "Predictive analytics," *Extending the Value of Your Data Warehousing Investment. TDWI Best Practices Report*, vol. 1, pp. 1-36, 2007.

[4] R. V. McCarthy, M. M. McCarthy, W. Ceccucci, L. Halawi, R. McCarthy, M. McCarthy, W. Ceccucci, and L. Halawi, *Applying predictive analytics*. Springer, 2022.

[5] A. H. Alsaeedi, S. M. Hadi, and Y. Alazzawi, "Adaptive Gamma and Color Correction for Enhancing Low-Light Images," *International Journal of Intelligent Engineering and Systems*, vol. 17, no. 4, pp. 188-201, 2024.

[6] A. H. Alsaeedi, H. H. R. Al-Mahmood, Z. F. Alnaseri, M. R. Aziz, D. Al-Shammary, A. Ibaida, and K. Ahmed, "Fractal feature selection model for enhancing high-dimensional biological problems," *BMC bioinformatics*, vol. 25, no. 1, p. 12, 2024.

[7] A. L. Albukhnefis, T. T. Al-Fatlawi, and A. H. Alsaeedi, "Image Segmentation Techniques: An In-Depth Review and Analysis," *Journal of Al-Qadisiyah for Computer Science and Mathematics*, vol. 16, no. 2, pp. Page 195-215, 2024.

[8] D. Al-Shammary, M. Radhi, A. H. AlSaeedi, A. M. Mahdi, A. Ibaida, and K. Ahmed, "Efficient ECG classification based on the probabilistic Kullback-Leibler divergence," *Informatics in Medicine Unlocked*, vol. 47, p. 101510, 2024.

[9] A. Y. Huang, O. H. Lu, J. C. Huang, C. Yin, and S. J. Yang, "Predicting students' academic performance by using educational big data and learning analytics: evaluation of classification methods and learning logs," *Interactive Learning Environments*, vol. 28, no. 2, pp. 206-230, 2020.

[10] A. H. Alsaeedi, M. A. Al-Sharqi, S. S. Alkafagi, R. R. Nuiaa, A. S. D. Alfoudi, S. Manickam, A. M. Mahdi, and A. M. Otebolaku, "Hybrid extend particle swarm optimization (EPSO) model for enhancing the performance of MANET routing protocols," *Journal of Al-Qadisiyah for computer science and mathematics*, vol. 15, no. 1, pp. Page 127-136, 2023.

[11] T. M. Renick, "Predictive analytics, academic advising, early alerts, and student success," *Big data on campus: Data analytics and decision making in higher education*, pp. 177-197, 2020.

[12] R. R. Nuiaa, S. A. A. A. Alsaidi, B. K. Mohammed, A. H. Alsaeedi, Z. A. A. Alyasseri, S. Manickam, and M. A. Hussain, "Enhanced PSO Algorithm for Detecting DRDoS Attacks on LDAP Servers," *International Journal of Intelligent Engineering & Systems*, vol. 16, no. 5, 2023.

[13] R. Al Ogaili, A. H. Alsaeedi, S. S. Alkafagi, and A. S. D. Alfoudi, "A critical review of Optimization MANET routing protocols," *Wasit Journal of Computer and Mathematics Science*, vol. 1, no. 4, pp. 44-54, 2022.

[14] E. Foster and R. Siddle, "The effectiveness of learning analytics for identifying at-risk students in higher education," *Assessment & Evaluation in Higher Education*, vol. 45, no. 6, pp. 842-854, 2020.

[15] R. R. Nuiaa, S. Manickam, A. H. Alsaeedi, and D. E. J. Al-Shammary, "Evolving Dynamic Fuzzy Clustering (EDFC) to Enhance DRDoS_DNS Attacks Detection Mechnism," *International Journal of Intelligent Engineering and Systems*, vol. 15, no. 1, pp. 509-519, 2022.

[16] R. R. Nuiaa, S. Manickam, and A. H. Alsaeedi, "A Comprehensive Review of DNS-based Distributed Reflection Denial of Service (DRDoS) Attacks: State-of-the-Art," *International Journal on Advanced Science, Engineering and Information*, vol. 6, no. 12, pp. 2452-2461, 2022.

[17] R. R. Nuiaa, S. Manickam, and A. S. D. Alfoudi, "Dynamic Evolving Cauchy Possibilistic Clustering Based on the Self-Similarity Principle (DECS) for Enhancing Intrusion Detection System," 2022.

[18] X. Bai, F. Zhang, J. Li, T. Guo, A. Aziz, A. Jin, and F. Xia, "Educational big data: Predictions, applications and challenges," *Big Data Research*, vol. 26, p. 100270, 2021.

[19] M. A. Ashaari, K. S. D. Singh, G. A. Abbasi, A. Amran, and F. J. Liebana-Cabanillas, "Big data analytics capability for improved performance of higher education institutions in the Era of IR 4.0: A multi-analytical SEM & ANN perspective," *Technological Forecasting and Social Change*, vol. 173, p. 121119, 2021.

[20] S. M. Ali, A. H. Alsaeedi, D. Al-Shammary, H. H. Alsaeedi, and H. W. Abid, "Efficient intelligent system for diagnosis pneumonia (SARSCOVID19) in X-ray images empowered with initial clustering," *Indones. J. Electr. Eng. Comput. Sci*, vol. 22, no. 1, pp. 241-251, 2021.

[21] A. S. Alfoudi, A. H. Alsaeedi, M. H. Abed, A. M. Otebolaku, and Y. Sajid Razooqi, "Palm Vein Identification Based on Hybrid Feature Selection Model," *International Journal of Intelligent Engineering and Systems*, vol. 14, no. 5, pp. 469-478, 2021.

[22] P. Alok, "Evaluation of Machine Learning Methods for Prediction Student Performance."

[23] D. S. Singh, P. Jain, A. Bhat, A. Vishwakarma, Hod, A. Proffessor, and P. Student, "Predictive Analytics for Student Performance and Academic Success: A Multi-Model Approach."

[24] R. Parkavi, P. Karthikeyan, S. Sujitha, and A. S. Abdullah, "Enhancing Educational Assessment: Predicting and Visualizing Student Performance using EDA and Machine Learning Techniques," *Journal of Engineering Education Transformations*, vol. 37, no. Special Issue 2, 2024.

[25] W. Sabbir, M. Abdullah-Al-Kafi, A. S. Afridi, M. S. Rahman, and M. Karmakar, "Improving Predictive Analytics for Student Dropout: A Comprehensive Analysis and Model Evaluation," in *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)*, 2024: IEEE, pp. 951-956.

[26] Z. Alamgir, H. Akram, S. Karim, and A. Wali, "Enhancing Student Performance Prediction via Educational Data Mining on Academic data," *Informatics in Education*, 2023, doi: 10.15388/infedu.2024.04.

[27] J. G. C. Ramírez, "Predictive Analytics in Education: Utilizing Machine Learning to Forecast Student Performance and Dropout Rates," *Asian American Research Letters Journal*, vol. 1, no. 5, 2024.

[28] J. Chukwunike and C. Cassia Anwar, "An Enhanced Student Engagement and Academic Performance Predictive System."

[29] M. Tastanova, R. Jilkibayeva, Z. Arkhabayeva, and S. Tripathi, "Enhancing Predictive Analytics in Education through Collaborative Filtering Algorithms in Knowledge Management," presented at the 2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM), 2023.

[30] A. C. Doctor, "A predictive model using machine learning algorithm in identifying students probability on passing semestral course," *arXiv preprint arXiv:2304.05565*, 2023.

[31] S. Albahli, "Efficient hyperparameter tuning for predicting student performance with Bayesian optimization," *Multimedia Tools and Applications*, vol. 83, no. 17, pp. 52711-52735, 2024.

[32] L. Ni, S. Wang, Z. Zhang, X. Li, X. Zheng, P. Denny, and J. Liu, "Enhancing student performance prediction on learnersourced questions with sgnn-llm synergy," in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024, vol. 38, no. 21, pp. 23232-23240.

[33] G. Ramaswami, T. Susnjak, and A. Mathrani, "Supporting students' academic performance using explainable machine learning with automated prescriptive analytics," *Big Data and Cognitive Computing*, vol. 6, no. 4, p. 105, 2022.

[34] S. Gaftandzhieva, A. Talukder, N. Gohain, S. Hussain, P. Theodorou, Y. K. Salal, and R. Doneva, "Exploring Online Activities to Predict the Final Grade of Student," *Mathematics*, vol. 10, no. 20, 2022, doi: 10.3390/math10203758.

[35] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learning Environments*, vol. 9, no. 1, 2022, doi: 10.1186/s40561-022-00192-z.

[36] M. H. B. Roslan and C. J. Chen, "Predicting students' performance in English and Mathematics using data mining techniques," *Educ Inf Technol (Dordr)*, vol. 28, no. 2, pp. 1427-1453, 2023, doi: 10.1007/s10639-022-11259-2.

[37] E. T. Khor, "A data mining approach using machine learning algorithms for early detection of low-performing students," *The International Journal of Information and Learning Technology*, vol. 39, no. 2, pp. 122-132, 2022, doi: 10.1108/ijilt-09-2021-0144.

[38] A. Bressane, M. Spalding, D. Zwirn, A. I. S. Loureiro, A. O. Bankole, R. G. Negri, I. de Brito Junior, J. K. S. Formiga, L. C. d. C. Medeiros, L. A. Pampuch Bortolozo, and R. Moruzzi, "Fuzzy Artificial Intelligence—Based Model Proposal to Forecast Student Performance and Retention Risk in Engineering Education: An Alternative for Handling with Small Data," *Sustainability*, vol. 14, no. 21, 2022, doi: 10.3390/su142114071.

[39] M. Adnan, A. Habib, J. Ashraf, S. Mussadiq, A. A. Raza, M. Abid, M. Bashir, and S. U. Khan, "Predicting at-Risk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models," *IEEE Access*, vol. 9, pp. 7519-7539, 2021, doi: 10.1109/access.2021.3049446.

[40] S. D. A. Bujang, A. Selamat, R. Ibrahim, O. Krejcar, E. Herrera-Viedma, H. Fujita, and N. A. M. Ghani, "Multiclass Prediction Model for Student Grade Prediction Using Machine Learning," *IEEE Access*, vol. 9, pp. 95608-95621, 2021, doi: 10.1109/access.2021.3093563.

[41] J. Kabathova and M. Drlik, "Towards Predicting Student's Dropout in University Courses Using Different Machine Learning Techniques," *Applied Sciences*, vol. 11, no. 7, 2021, doi: 10.3390/app11073130.

[42] M. N. Yakubu and A. M. Abubakar, "Applying machine learning approach to predict students' performance in higher educational institutions," *Kybernetes*, vol. 51, no. 2, pp. 916-934, 2021, doi: 10.1108/k-12-2020-0865.

[43] J. Malini and Y. Kalpana, "Investigation of factors affecting student performance evaluation using education materials data mining technique," *Materials Today: Proceedings*, vol. 47, pp. 6105-6110, 2021, doi: 10.1016/j.matpr.2021.05.026.

[44] R. Bertolini, S. J. Finch, and R. H. Nehm, "Enhancing data pipelines for forecasting student performance: integrating feature selection with cross-validation," *International Journal of Educational Technology in Higher Education*, vol. 18, pp. 1-23, 2021.

[45] I. El Guabassi, Z. Bousalem, R. Marah, and A. Qazdar, "Comparative Analysis of Supervised Machine Learning Algorithms to Build a Predictive Model for Evaluating Students' Performance," *International Journal of Online and Biomedical Engineering (iJOE)*, vol. 17, no. 02, pp. 90-105, 2021, doi: 10.3991/ijoe.v17i02.20025.

[46] R. Hasan, S. Palaniappan, S. Mahmood, A. Abbas, K. U. Sarker, and M. U. Sattar, "Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques," *Applied Sciences*, vol. 10, no. 11, 2020, doi: 10.3390/app10113894.

[47] A. S. Rayasam, "Predicting at-risk students from disparate sources of institutional data," Massachusetts Institute of Technology, 2020.

[48] N. Yılmaz and B. Sekeroglu, "Student Performance Classification Using Artificial Intelligence Techniques," in *10th International Conference on Theory and Application of Soft Computing, Computing with Words and Perceptions - ICSCCW-2019*, (Advances in Intelligent Systems and Computing, 2020, ch. Chapter 76, pp. 596-603.

[49] M. R. Alzahrani, "Predicting Student Performance Using Ensemble Models and Learning Analytics Techniques," 2024.

[50] A. Al-Ameri, W. Al-Shammari, A. Castiglione, M. Nappi, C. Pero, and M. Umer, "Student Academic Success Prediction Using Learning Management Multimedia Data With Convoluted Features and Ensemble Model," *Journal of Data and Information Quality*, 2024, doi: 10.1145/3687268.

[51] A. A. Nafea, M. Mishlish, A. M. S. Shaban, M. M. Al-Ani, K. M. A. Alheeti, and H. J. Mohammed, "Enhancing Student's Performance Classification Using Ensemble Modeling," *Iraqi Journal For Computer Science and Mathematics*, vol. 4, no. 4, pp. 204-214, 2023, doi: 10.52866/ijcsm.2023.04.04.016.

[52] Y. S. Balcioğlu and M. Artar, "Predicting academic performance of students with machine learning," *Information Development*, 2023, doi: 10.1177/02666669231213023.

[53] M. M. Hussain, S. Akbar, S. A. Hassan, M. W. Aziz, and F. Urooj, "Prediction of Student's Academic Performance through Data Mining Approach," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 241-251, 2024, doi: 10.33093/jiwe.2024.3.1.16.

[54] K. Sharada, "Student Psychology Teaching Learning Optimisation-Based Deep Long Short-Term Memory for Predicting Student Performance," *Journal of Information & Knowledge Management*, vol. 23, no. 03, 2024, doi: 10.1142/s0219649224500400.

[55] F. A. Al-Azazi and M. Ghurab, "ANN-LSTM: A deep learning model for early student performance prediction in MOOC," *Heliyon*, vol. 9, no. 4, p. e15382, Apr 2023, doi: 10.1016/j.heliyon.2023.e15382.

[56] R. Y.-Y. Chan, C. M. V. Wong, and Y. N. Yum, "Predicting Behavior Change in Students With Special Education Needs Using Multimodal Learning Analytics," *IEEE Access*, vol. 11, pp. 63238-63251, 2023, doi: 10.1109/access.2023.3288695.

[57] D. A. Shafiq, M. Marjani, R. A. A. Habeeb, and D. Asirvatham, "A Conceptual Predictive Analytics Model for the Identification of at-risk students in VLE using Machine Learning Techniques," presented at the 2022 14th International Conference on Mathematics, Actuarial Science, Computer Science and Statistics (MACS), 2022.

[58] T. Tao, C. Sun, Z. Wu, J. Yang, and J. Wang, "Deep Neural Network-Based Prediction and Early Warning of Student Grades and Recommendations for Similar Learning Approaches," *Applied Sciences*, vol. 12, no. 15, 2022, doi: 10.3390/app12157733.

[59] C. Liu, H. Wang, Y. Du, and Z. Yuan, "A Predictive Model for Student Achievement Using Spiking Neural Networks Based on Educational Data," *Applied Sciences,* vol. 12, no. 8, 2022, doi: 10.3390/app12083841.

[60] C. F. Rodríguez-Hernández, M. Musso, E. Kyndt, and E. Cascallar, "Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation," *Computers and Education: Artificial Intelligence,* vol. 2, 2021, doi: 10.1016/j.caeai.2021.100018.

[61] A. Nabil, M. Seyam, and A. Abou-Elfetouh, "Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks," *IEEE Access,* vol. 9, pp. 140731-140746, 2021, doi: 10.1109/access.2021.3119596.

[62] H. Waheed, S.-U. Hassan, R. Nawaz, N. R. Aljohani, G. Chen, and D. Gasevic, "Early prediction of learners at risk in self-paced education: A neural network approach," *Expert Systems with Applications,* vol. 213, p. 118868, 2023.

[63] R. C. Raga and J. D. Raga, "Early prediction of student performance in blended learning courses using deep neural networks," in *2019 International Symposium on Educational Technology (ISET)*, 2019: IEEE, pp. 39-43.

[64] M. Tsiakmaki, G. Kostopoulos, S. Kotsiantis, and O. Ragos, "Transfer Learning from Deep Neural Networks for Predicting Student Performance," *Applied Sciences,* vol. 10, no. 6, 2020, doi: 10.3390/app10062145.

[65] M. Karlık and B. Karlık, "Prediction of student's performance with deep neural networks," *Int. J. of Artificial Intel. and Expert Sys.(IJAE),* vol. 9, no. 2, pp. 39-47, 2020.

[66] S. Hussain, S. Gaftandzhieva, M. Maniruzzaman, R. Doneva, and Z. F. Muhsin, "Regression analysis of student academic performance using deep learning," *Education and Information Technologies,* vol. 26, no. 1, pp. 783-798, 2020, doi: 10.1007/s10639-020-10241-0.

[67] A. Almasri, R. S. Alkhawaldeh, and E. Çelebi, "Clustering-Based EMT Model for Predicting Student Performance," *Arabian Journal for Science and Engineering,* vol. 45, no. 12, pp. 10067-10078, 2020, doi: 10.1007/s13369-020-04578-4.

[68] T. Omar, A. Alzahrani, and M. Zohdy, "Clustering Approach for Analyzing the Student's Efficiency and Performance Based on Data," *Journal of Data Analysis and Information Processing,* vol. 08, no. 03, pp. 171-182, 2020, doi: 10.4236/jdaip.2020.83010.

[69] Y. L. Goh, Y. H. Goh, C.-C. Yip, C. H. Ting, R. L. Leh Bin, and K. P. Chen, "Students' Academic Performance Analysis by K-Means Clustering for Investigating Students' Health Conditions within Clusters," *Annals of Tropical Medicine and Public Health,* vol. 23, no. 13, 2020, doi: 10.36295/asro.2020.231332.

[70] E. Ahmed, "Student Performance Prediction Using Machine Learning Algorithms," *Applied Computational Intelligence and Soft Computing,* vol. 2024, no. 1, p. 4067721, 2024.

[71] K. Vimarsha, S. S. Prakash, K. Krinkin, and Y. A. Shichkina, "Student Performance Prediction: A Co-Evolutionary Hybrid Intelligence model," *Procedia Computer Science,* vol. 235, pp. 436-446, 2024.

[72] G. Al-Tameemi, J. Xue, I. H. Ali, and S. Ajit, "A Hybrid Machine Learning Approach for Predicting Student Performance Using Multi-class Educational Datasets," *Procedia Computer Science,* vol. 238, pp. 888-895, 2024.

[73] A. Alshanqiti and A. Namoun, "Predicting Student Performance and Its Influential Factors Using Hybrid Regression and Multi-Label Classification," *IEEE Access,* vol. 8, pp. 203827-203844, 2020, doi: 10.1109/access.2020.3036572.

[74] E. Evangelista, "A Hybrid Machine Learning Framework for Predicting Students' Performance in Virtual Learning Environment," *International Journal of Emerging Technologies in Learning (iJET),* vol. 16, no. 24, pp. 255-272, 2021, doi: 10.3991/ijet.v16i24.26151.

[75] A. A. Alharbi and J. Allohibi, "A new hybrid classification algorithm for predicting student performance," *AIMS Mathematics,* vol. 9, no. 7, pp. 18308-18323, 2024, doi: 10.3934/math.2024893.

[76] J. Kuzilek, M. Hlosta, and Z. Zdrahal, "Open university learning analytics dataset," *Scientific data,* vol. 4, no. 1, pp. 1-8, 2017.

[77] K.-E. l. system. *Students' Academic Performance Dataset*. [Online]. Available: https://www.kaggle.com/datasets/aljarah/xAPI-Edu-Data

[78] P. Cortez, "Student Performance [Dataset]," *UCI Machine Learning Repository,* 2008, doi: https://doi.org/10.24432/C5TG7T.

[79] P. Samarati and L. Sweeney, "Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression," 1998.

[80] S. Barocas, M. Hardt, and A. Narayanan, "Fairness and Machine Learning Limitations and Opportunities," 2018.